

1/5/1
DIALOG(R)File 351:Derwent WPI
(c) 2000 Derwent Info Ltd. All rts. reserv.

09/700311
529 Rec'd PCT/PTC 13 NOV2000

009298580 **Image available**
WPI Acc No: 1992-425989/*199252*
XRPX Acc No: N92-325030

**Speech synthesis technique - by decomposing signal orthogonally on
Daubechies wavelet base using coeffs. stored and selected w.r.t.
synthesising algorithm and reconstituting from coeffs.**

Patent Assignee: SEXTANT AVIONIQUE (SEXT-N)
Inventor: GULLI C
Number of Countries: 009 Number of Patents: 005
Patent Family:

| Patent No | Kind | Date | Applicat No | Kind | Date | Week |
|------------|------|----------|-------------|------|----------|-------------|
| EP 519802 | A1 | 19921223 | EP 92401663 | A | 19920616 | 199252 B |
| WO 9222890 | A1 | 19921223 | WO 92FR538 | A | 19920616 | 199302 |
| FR 2678103 | A1 | 19921224 | FR 917424 | A | 19910618 | 199308 |
| JP 6503186 | W | 19940407 | WO 92FR538 | A | 19920616 | 199419 |
| | | | JP 93500767 | A | 19920616 | |
| US 5826232 | A | 19981020 | WO 92FR538 | A | 19920616 | 199849 |
| | | | US 93972486 | A | 19930218 | |

Priority Applications (No Type Date): FR 917424 A 19910618
Cited Patents: 06Jnl.Ref; FR 2648567; US 4599567; US 4974187

Patent Details:

| Patent No | Kind | Lan | Pg | Main IPC | Filing Notes |
|--|------|-----|----|-------------|----------------------------|
| EP 519802 | A1 | F | 12 | G10L-005/02 | |
| Designated States (Regional): BE DE ES FR GB NL SE | | | | | |
| WO 9222890 | A1 | F | 20 | G10L-005/02 | |
| Designated States (National): JP US | | | | | |
| JP 6503186 | W | | | G10L-009/16 | Based on patent WO 9222890 |
| US 5826232 | A | | | G10L-009/00 | Based on patent WO 9222890 |
| FR 2678103 | A1 | | | G10L-009/06 | |

Abstract (Basic): EP 519802 A

The voice synthesis technique includes initial digitising of a natural voice signal. This signal is then cut using a base of orthogonal wavelets with a compact support. The coefficients representing the voice signal are then stored.

The voice signal may then be reconstituted using these coefficients in a process of filtering, interpolation and low frequency amplification. The choice of regularity of the synthesis wavelets is determined by a preliminary analysis of the speech frames.

USE/ADVANTAGE - E.g. for voice control system for vehicle, or industrial process. Simple processing system of reduced cost using widely available components. Reduced memory for message storage.

Dwg.3/9

Title Terms: SPEECH; SYNTHESIS; TECHNIQUE; DECOMPOSE; SIGNAL; ORTHOGONAL;
BASE; COEFFICIENT; STORAGE; SELECT; SYNTHESIS; ALGORITHM; RECONSTITUTED;
COEFFICIENT

Derwent Class: P86; W04

International Patent Class (Main): G10L-005/02; G10L-009/00; G10L-009/06;
G10L-009/16

International Patent Class (Additional): G10L-009/18

File Segment: EPI; EngPI



Europäisches Patentamt
European Patent Office
Office européen des brevets



Numéro de publication : **0 519 802 A1**

12

DEMANDE DE BREVET EUROPEEN

21 Numéro de dépôt : 92401663.7

51 Int. Cl.⁵ : **G10L 5/02, G10L 9/16**

22 Date de dépôt : 16.06.92

30 Priorité : 18.06.91 FR 9107424

43 Date de publication de la demande :
23.12.92 Bulletin 92/52

84 Etats contractants désignés :
BE DE ES FR GB NL SE

71 Demandeur : **SEXTANT AVIONIQUE**
5,7, rue Jeanne Braconnier Parc Tertiaire
F-92360 Meudon-la-Forêt (FR)

72 Inventeur : **Gulli, Christian**
THOMSON-CSF, SCPI, Cédex 67
F-92045 Paris la Défense (FR)

74 Mandataire : **Chaverneff, Vladimir et al**
THOMSON-CSF SCPI B.P. 329
F-92402 COURBEVOIE CEDEX (FR)

54 Procédé de synthèse vocale à l'aide d'ondelettes.

57 La synthèse vocale de l'invention consiste à analyser un signal vocal par décomposition orthogonale sur une base d'ondelettes à support compact, de préférence des ondelettes de Daubechies. La synthèse est effectuée à partir des coefficients stockés et sélectionnés lors de l'analyse, selon le même algorithme que celui utilisé pour l'analyse.

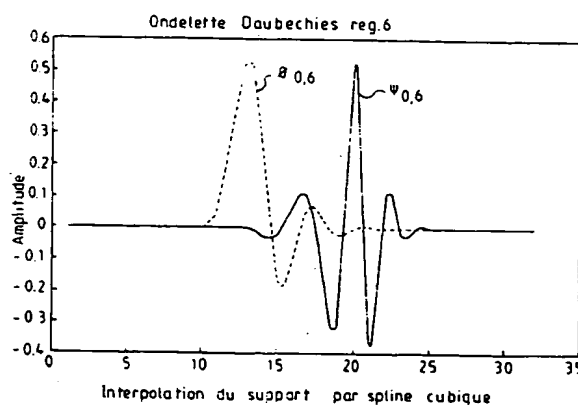


FIG.8

EP 0 519 802 A1

La présente invention se rapporte à un procédé de synthèse vocale.

Parmi les nombreux domaines d'application de la synthèse vocale, certains, tels que les appareils à commande interactive (commande de véhicules, de processus industriels, ...) ne nécessitent que la synthèse de messages simples (mots isolés ou phases prédéterminées). Dans de telles applications, on recherche à minimiser le coût du dispositif de synthèse vocale. La diminution du coût peut être essentiellement obtenue par l'emploi de circuits de grande diffusion et par la diminution de la capacité de mémoire nécessaire au stockage des messages.

Afin de réduire cette capacité de mémoire, l'art connu fait appel à différents types de codage. Parmi les codages les plus employés, on connaît le codage temporel qui associe à l'amplitude du signal un code binaire à des instants discrets, et plus précisément, on mémorise plutôt la différence entre le signal et sa composante prédictible (codage différentiel). On fait également appel au codage de la parole par analyse et synthèse, selon lequel on ne stocke que très peu de paramètres significatifs (dispositifs dits : "vocodeur à canaux" ou "vocodeur à prédiction linéaire"). On connaît enfin un procédé qui résulte de l'association des deux procédés précités : "vocodeur prédictif adaptatif" ou "à excitation vocale", en particulier le codage en sous-bandes.

Dans le cas du codage en sous-bandes, qui est un codage dans le domaine fréquentiel, on fractionne le spectre du signal à coder en un certain nombre de sous-bandes de largeur B_k (égales entre elles ou non). Chaque sous-bande (d'indice k) est ensuite rééchantillonnée à la fréquence de Shannon, soit $2B_k$. Les signaux issus de chaque filtre de sous-bande sont quantifiés différemment en fonction de la fréquence, à savoir quantification fine pour le fondamental et les formants, et quantification grossière dans les plages où l'énergie est faible. L'opération inverse est réalisée pour reconstruire le signal.

Avant stockage et transmission, les signaux sont codés par exemple selon une loi de codage MIC (modulation par impulsions et codage) normalisée à 64 kbits/s (signal échantillonné à 8 kHz sur 8 bits dans la bande 300-3600 Hz et compressé selon une loi logarithmique). Le codage MICDA (MIC différentiel adaptatif), à un débit de 32 kbits/s (8 kHz sur 4 bits), tend à se généraliser.

On a représenté en figure 1 le schéma de principe d'un dispositif 1 de codage à deux sous-bandes. Le signal de parole x est filtré par deux filtres F_1 , F_2 (de réponses impulsionnelles h_1 , h_2). Chacune des deux sous-bandes en sortie de F_1 , F_2 est décimée par 2 (suppression d'un échantillon sur 2) par les circuits 2, 3 respectivement, puis codée (4), par exemple en MICDA et stockée (ou transmise). A la lecture (ou réception), la reconstitution du signal de parole se fait par décodage (5, 6) puis filtrage dans des interpola-

teurs (7, 8) identiques à ceux de la bande d'analyse correspondante et sommation (9) des deux sous-bandes décodées. Les filtres F_1 et F_2 sont des filtres FIR (à réponse impulsionnelle finie) à phase linéaire et satisfont aux conditions suivantes.

$$h_2(n) = (-1)^n h_1(n) \\ |H_1(e^{j\theta})|^2 + |H_2(e^{j\theta})|^2 = 1$$

On a représenté en figure 2 le gabarit de ces filtres.

Le principe du codage en sous-bandes consiste à filtrer le signal de parole par un banc de filtres, puis à sous-échantillonner les signaux de sortie de ces filtres. A la réception, la reconstitution se fait par addition de chaque sous-bande décodée, interpolée par un filtre identique à celui de la bande d'analyse correspondante. Ce type de codage a d'abord été implanté à partir de filtres disjoints et contigus à réponse impulsionnelle finie. Il a ensuite été étendu grâce à l'emploi de filtres miroirs en quadrature, permettant une reconstitution quasi parfaite du signal initial en l'absence d'erreur sur la quantification.

Il existe deux grandes familles de procédés pour synthétiser les filtres qui décomposent le signal de parole :

- soit on scinde l'entrée en deux bandes par un filtre optimisé, et on renouvelle l'algorithme pour chaque bande ;
- soit on déplace sur l'axe fréquentiel un gabarit de filtre passe-bande. Dans ce cas, le filtre de base est de réponse $h(n)$ et de largeur de bande $\pi/2M$ (M étant le nombre de sous-bandes). Par déplacement, on obtient :

$$h_i(n) = h(n) \cdot \cos(n\pi(2i+1)/2M)$$

π étant la demi-fréquence d'échantillonnage normalisée. Le problème de repliement des gabarits lors du sous-échantillonnage peut être compensé par un terme de phase dans la fonction cosinus de déphasage.

Le filtre demi-bande, dont le gabarit est représenté en figure 2, est classiquement un filtre linéaire dont la fonction de transfert vaut $1/2$ à $f_e/4$ (f_e = fréquence d'échantillonnage) et est antisymétrique par rapport à ce point, c'est-à-dire que l'on a :

$$H_1\left(f_e/4 + f\right) = 1 - H\left(f_e/4 - f\right)$$

Les coefficients $h(n)$ sont nuls pour n pair, sauf h_0 . Le gabarit est défini par l'ondulation en bandes passante et coupée, et par Δf qui représente la largeur de la bande de transition. Le nombre N de coefficients du filtre en fonction du gabarit désiré est donné par la relation approchée :

$$N \approx \frac{2}{3} \log \left(\frac{1}{10\delta^2} \right) \frac{f_e}{\Delta f}$$

dans laquelle $\delta = \delta_1 = \delta_2$ représente l'ondulation dans les bandes passante et coupée. La réduction ou l'élé-

vation de la fréquence d'échantillonnage sont obtenues par la mise en cascade de P filtres demi-bande. La fréquence intermédiaire f_i est un sous-multiple de la fréquence d'échantillonnage dans un rapport deux : $f_e = 2^P \cdot f_i$.

Il existe également des dispositifs procédant à l'analyse multirésolution du signal de parole, et comportant essentiellement un filtre discret et un circuit de "décimation" (suppression d'un échantillon sur deux). On connaît également ("Traitement de Signal", vol 7, n° 2, 1990), pour la compression numérique d'image, un algorithme rapide mettant en oeuvre une transformée en ondelettes, mais cet algorithme ne convient que pour des images (seule la composante HF est conservée).

Les dispositifs connus sont soit trop rudimentaires, et ne permettent pas d'obtenir à la restitution un signal de parole suffisamment intelligible, soit trop complexes et donc onéreux.

La présente invention a pour objet un procédé de synthèse vocale qui permette de synthétiser le plus simplement possible des signaux de parole et ne fasse appel, pour sa mise en oeuvre, qu'à des circuits existants et peu onéreux.

Le procédé de l'invention consiste à numériser un signal vocal, à découper ce signal numérisé dans une base orthogonale d'ondelettes à support compact, à stocker les coefficients représentatifs du signal vocal, et, à la restitution, à reconstituer le signal vocal par filtrage, interpolation et amplification basse fréquence.

L'invention sera mieux comprise à la lecture de la description détaillée d'un mode de réalisation, pris à titre d'exemple non limitatif et illustré par le dessin annexé, sur lequel :

- la figure 1, déjà décrite ci-dessus, est un bloc-diagramme d'un système de codage connu ;
- la figure 2 est un gabarit de filtre demi-bande utilisable dans le système de la figure 1 ;
- la figure 3 est un bloc-diagramme d'un système de synthèse mettant en oeuvre le procédé conforme à l'invention ;
- la figure 4 est un bloc-diagramme du dispositif d'analyse du système de la figure 3 ;
- la figure 5 est un diagramme illustrant l'algorithme de décomposition de l'invention ;
- la figure 6 est un diagramme illustrant l'algorithme de reconstruction de l'invention ;
- la figure 7 est un bloc-diagramme simplifié d'un dispositif de synthèse vocale mettant en oeuvre le procédé de l'invention ;
- la figure 8 est un chronogramme d'une fonction d'échelle et d'une ondelette utilisées par l'invention ; et
- la figure 9 est un schéma d'un dispositif de synthèse mettant en oeuvre le procédé conforme à l'invention.

Le synthétiseur de messages vocaux décrit ci-dessous comporte deux parties principales : une par-

tie 14 d'analyse et une partie 15 de synthèse vocale (figure 3).

Dans la partie 14, les signaux de la source 16 (par exemple un microphone) sont quantifiés, puis analysés en 17 et codés en 18. Les critères pertinents qui en résultent sont stockés en 19 (par exemple des mémoires de type EEPROM). Toutes ces opérations sont, dans le cas présent, réalisées en laboratoire.

Dans la deuxième partie, qui comporte le dispositif de stockage 19, un dispositif 20 assure la reconstitution du signal à partir des coefficients sélectionnés et stockés (en 19), le signal reconstitué est envoyé à un amplificateur 21 muni d'un haut-parleur.

Selon l'invention, on met en oeuvre, pour le codage et la reconstitution, un algorithme qui décompose le signal vocal dans une base orthogonale d'ondelettes à support compact. Ces ondelettes sont par exemple des ondelettes de Daubechies (voir figure 8). Seuls sont stockés les coefficients jugés représentatifs du signal vocal de départ et assurant une parfaite intelligibilité du message reconstitué, ce qui limite fortement le débit des signaux à stocker.

L'organigramme de la figure 4 illustre la procédure d'analyse vocale conforme à l'invention.

Les signaux basse fréquence produits par une source de signaux basse fréquence 22 (capteur acoustique, moyen de stockage magnétique, ...) sont numérisés (23), par exemple sur 16 bits, par exemple à l'aide d'un convertisseur "flash" ou d'un convertisseur à approximations successives (dont le temps de conversion est de l'ordre de 60 μ s ou moins) à une fréquence d'échantillonnage, qui est par exemple de 10 kHz. Le signal échantillonné est ensuite découpé en trames de, par exemple, 128 points (durée d'une trame : 12,8 ms). Selon un autre exemple, on peut mettre en oeuvre des trames de 256 points, sans préjudice notable pour la qualité de la restitution. Ensuite, on procède à l'analyse (24), qui constitue une étape essentielle de l'invention. Cette analyse consiste en particulier à décomposer le signal numérisé sur une base orthogonale d'ondelettes à support compact, et fait appel à des filtres dont la réponse impulsionnelle peut être symétrique ou non. Dans le cas où cette réponse est symétrique, on limite le stockage des coefficients extrêmes (responsables des effets de bords) à un seul côté du signal, l'autre côté étant déduit par symétrie (la périodicité des filtres est implicite par construction).

A partir des 128 points initiaux, on obtient donc par cette décomposition 128 combinaisons linéaires indépendantes de la base d'observation. La régularité de l'onde, qui conditionne la forme du filtre de décomposition est un des deux paramètres majeurs de la décomposition (avec le niveau de décomposition, qui conditionne la largeur du filtre). Parmi ces 128 combinaisons, on en conserve par exemple 32 (estimées être les plus significatives) qui sont codées (25), dans le cas présent sur 8 bits, ce qui donne un débit de va-

leurs à stocker de 20 kbits/s. La sélection de 16 coefficients codés sur 16 bits ne modifierait pas le débit des valeurs à stocker, mais diminuerait la qualité du signal restitué.

On notera que l'analyse par dilatation de l'échelle des temps (voir fonction d'échelle, en trait interrompu, en figure 8) est réalisée non pas en dilatant les ondelettes d'analyse, mais en sous-échantillonnant d'un facteur 2^p le signal à analyser. Il en résulte, pour une décomposition à un niveau p , $(p+1)$ jeux de coefficients. De plus, la projection sur une base orthogonale (à nombre de points = $N/2 + N/4 + \dots + N/2^{p+1}$) n'induit ni perte ni redondance d'informations. La repré-

sentation en ondelettes devient $(S_j, D_j)_{0 \leq j \leq J}$ où S_j est l'approximation du signal à la résolution 2^j et les D_j correspondent aux détails de résolution 2^j .

Les paramètres ayant été codés (25), on procède toujours en laboratoire, avant de les stocker, à une évaluation (26), en réalisant la synthèse, comme décrit ci-dessous. Si (en 27) la qualité de la restitution du signal vocal est mauvaise, on modifie (28) le choix des paramètres résultant de l'analyse (24), et on les code (25) pour une nouvelle évaluation (25). Si cette qualité est jugée bonne, on met en forme les trames de paramètres (29) et on transmet celles-ci, par exemple via une liaison série RS422 (30), aux moyens de stockage.

On a illustré en figure 5 la mise en oeuvre de l'algorithme de décomposition selon l'invention.

Les différentes composantes S_0 à S_j sont traitées chacune de la même façon : convolution avec les $(j+1)$ filtres \tilde{c}_j (31.0 à 31.j) et leurs $(j+1)$ miroirs res-

pectifs (32.0 à 32.j) \tilde{H} et décimation par 2 (respectivement 32.0 à 32.j et 34.0 à 34.j).

Pour une régularité n , le support du filtre comporte $2.n$ valeurs. A partir des N coefficients de départ, on a pour $n=1$ 2 fois $N/2$ coefficients, pour $N=2$, 4 fois $N/4$ coefficients, etc..., mais on n'en stocke que $N/2n$. Si l'on prend par exemple $n=6$, on met en oeuvre une convolution sur 12 points. Cette valeur implique que la convolution est réalisée dans le domaine temporel. Cependant, pour une régularité supérieure à environ 16, il est préférable, du point de vue du temps de calcul du processeur d'analyse, de substituer à la convolution une multiplication dans l'espace fréquentiel dual (ce qui revient à une convolution sectionnée).

Le codage des paramètres (en 25) peut être réalisé soit à partir d'histogrammes locaux, soit, de manière plus simple, par une quantification liée à un niveau énergétique fixé par avance.

La phase d'évaluation (26) consiste à écouter le message reconstitué, et, le cas échéant, si l'écoute n'est pas jugée satisfaisante, à modifier (28) les paramètres à stocker. Cette reconstitution se fait, comme décrit ci-dessous en détail, par conversion numérique/analogique, filtrage passe-bas de lissage

et amplification basse fréquence. Lorsque la qualité du message reconstitué est jugée satisfaisante, on met en forme les coefficients (29) et on les charge (30) dans une mémoire appropriée. La mise en forme consiste essentiellement à formater les données, à produire les adresses correspondantes et à séquencer les trames successives de données.

On a illustré en figure 6 l'algorithme de synthèse vocale proprement dit mettant en oeuvre le procédé de l'invention, qui constitue un moyen autonome de génération de messages, distinct du dispositif de synthèse de laboratoire, mentionné ci-dessus, ayant servi pour l'évaluation du choix des paramètres. Cet algorithme de synthèse vocale reconstitue le signal d'origine en procédant par interpolation (35.0 à 35.j pour S_0 à S_j et 36.0 à 36.j pour D_0 à D_j), filtrage (37.0 à 37.j et 38.0 à 38.j respectivement), addition (39.0 à 39.j), multiplication (40.0 à 40.j) et amplification basse fréquence. En effet, à partir de la décomposition en échelle-ondelette au niveau p (typiquement $p=2$ à 3), il est possible de reconstruire la décomposition au niveau $(p-1)$. Il suffit pour cela d'insérer des valeurs nulles entre chaque valeur de la décomposition au niveau p , puis de convoluer avec les fonctions ondelettes et échelles inverses selon l'algorithme de reconstruction détaillé ci-dessus.

Les ondelettes de Daubechies, que l'invention utilise préférentiellement, sont des ondelettes à support compact, qui minimisent de ce fait le nombre de points de leur réponse impulsionnelle, donc de la convolution.

Les filtres de décomposition sont identiques à ceux de reconstruction, mais ils ne sont pas symétriques, ce qui oblige à mémoriser les coefficients dus aux effets de bords au début et à la fin de la trame de coefficients à mémoriser. On peut contourner ce problème en utilisant des ondelettes bi-orthogonales, ce qui oblige alors à utiliser des filtres de reconstruction différents de ceux de décomposition, mais leur réponse étant symétrique, seuls sont stockés les coefficients d'un seul côté.

On a représenté en figure 7 le schéma simplifié d'un dispositif de synthèse vocale mettant en oeuvre le procédé conforme à l'invention. Les coefficients des filtres de reconstruction sont stockés dans une mémoire 41 et utilisés par un calculateur spécialisé ou un microprocesseur 42 qui reconstruit le signal vocal sous la commande de l'algorithme de reconstruction décrit ci-dessus et mémorisé dans sa mémoire de programme 43 avec les valeurs des réponses impulsionnelles des différents filtres de reconstruction. Les valeurs numériques du signal reconstruit sont converties en analogique par le convertisseur 44 qui est suivi d'un amplificateur 45 à filtre analogique passe-bas (à fréquence de coupure de 4 kHz par exemple) et à commande de gain 46. La sortie de l'amplificateur 45 est reliée à un haut-parleur 47. L'amplificateur comporte avantageusement une sortie haute

impédance 48 qui peut être reliée à un dispositif d'enregistrement approprié. Le microprocesseur 42 est par ailleurs relié à une entrée 49 (par exemple entrée série RS232 ou RS422) par laquelle il reçoit des demandes de synthèse de messages vocaux. Ces demandes peuvent provenir de circuits d'alarme.

Sur le schéma détaillé du dispositif de synthèse vocale de la figure 9, on a représenté le processeur 50 avec son bus d'adresses 51, son bus de données 52 et son bus de commande 53, qui est relié en particulier à un séquenceur logique 54. Le séquenceur est relié à une interface série d'entrée 55 et à une interface série de sortie 56, et via un circuit d'optoisolation 57 à un dispositif de commande de synthèse de messages (non représenté), qui lui envoie les adresses des messages à synthétiser. Une mémoire de programme 58 est reliée au trois bus 51 à 53. Les coefficients sont stockés dans une mémoire 59 reliée directement au bus d'adresses et au séquenceur 54 et reliée via une porte trois états 60 au bus de données, la porte 60 étant commandée par le séquenceur 54.

Les bus 51 à 53 peuvent être reliés à un connecteur extérieur pour télécharger des coefficients ou modifier le programme de reconstruction, pour effectuer des tests ou des travaux de maintenance.

Le séquenceur 54 est relié à un convertisseur numérique/analogique 61 suivi d'un filtre passe-bas 62 et d'un amplificateur basse fréquence 63 dont le gain peut être ajusté par un potentiomètre 64. L'amplificateur 63 est relié à un ou plusieurs haut-parleurs 65 et à une borne de sortie haute impédance 66.

Le traitement des effets de bords est rendu indispensable lorsque l'on utilise un niveau de décomposition important. Il peut être obtenu en rendant artificiellement impaires les trames de parole, en ajoutant d'un côté d'une trame de parole ou des deux côtés une copie d'une partie de cette trame, par exemple pour une trame de 256 points, on rajoute d'un côté ou des deux 128 points.

Il est possible d'adopter une modélisation autorégressive de la trame (25,6 ms) de parole voisée pour étendre artificiellement sa durée par une extrapolation temporelle.

Le traitement de synthèse qui est décrit ci-dessus par blocs peut être mis en oeuvre par N filtres disjoints en cascade (type vocodeur). Ce procédé limite les effets de bords dus au rafraichissement des valeurs du filtrage, mais pénalise le processeur puisque l'on n'utilise pas alors les optimisations décrites lors de la décomposition dyadique.

La base orthogonale choisie est à support compact, ce qui optimise le temps de calcul de la convolution du filtrage. Les coefficients sont réels, ce qui permet une interprétation facile du module et du signe, et qui relaxe les contraintes liées à l'exploitation physique du modulo 2π (lorsque la base est complexe). Lorsque le nombre de points utilisés est

inférieur à environ 30, on effectue une convolution temporelle. On peut faire appel à plusieurs bases orthogonales, avec des régularités différentes.

- la décomposition n'est pas établie à un niveau donné, mais chaque filtre est adapté en largeur (par exemple niveau de décomposition oblique : analyse à Q

$$Q \left\{ \frac{\Delta f}{f} \right\}$$

constant) grâce au niveau variable en fonction de l'optimisation liée à la parole. On peut par exemple effectuer un découpage plus fin autour de 800 Hz ;

- le choix de la régularité de l'ondelette de synthèse peut par exemple être déterminé par une analyse préliminaire des trames de parole (par "ondelette de voisement" qui est par exemple une ondelette moyenne déterminée à partir des trois classes de voisement ou la dérivée troisième d'une gaussienne) ;

- trame voisée (structure harmonique) : régularité 6 à 10 environ ;

- trame non voisée (plosives, fricatives) : faible régularité (1 à 6) ;

- le réagencement des coefficients d'ondelette (résultat du produit scalaire) en fonction de leur position fréquentielle permet de traiter plus facilement l'analyse temps-échelle et de la voir comme une analyse temps-fréquence ;

- une quantification vectorielle permet d'optimiser le débit en adaptant le codage en fonction du rang fréquentiel et de l'énergie à coder. Quelle que soit la méthode employée (par exemple dichotomie), la finalité reste toujours l'élaboration d'un "codebook" multirésolution (un "codebook" étant un ensemble de vecteurs qui comportent toutes les "classes" ou vecteurs caractérisant le barycentre de nuages de points). In fine on essaye de choisir une distorsion minimale (erreur quadratique faible) aussi peu pénalisante que possible ;

- le nombre de bits de codage d'un vecteur du code book est fonction de l'énergie traitée (nombre important pour le fondamental, faible pour les fréquences extrêmes).

Revendications

1. Procédé de synthèse vocale, caractérisé par le fait qu'il consiste à numériser un signal vocal, à découper ce signal numérisé dans au moins une base orthogonale d'ondelettes à support compact, à stocker les coefficients représentatifs du signal vocal, et, à la restitution, à reconstituer

le signal vocal par filtrage, interpolation et amplification basse fréquence.

2. Procédé selon la revendication 1, caractérisé par le fait que les coefficients sont réels. 5
3. Procédé selon la revendication 1 ou 2, caractérisé par le fait que le choix de la régularité de l'ondelette de synthèse est déterminé par une analyse préliminaire des trames de parole. 10
4. Procédé selon l'une des revendications précédentes, caractérisé par le fait que pour des trames de parole voisée, la régularité de l'ondelette de synthèse est d'environ 6 à 10. 15
5. Procédé selon l'une des revendications 1 à 3, caractérisé par le fait que pour des trames de parole non voisée, la régularité de l'ondelette de synthèse est de 1 à 6. 20
6. Procédé selon l'une des revendications précédentes, caractérisé par le fait que pour traiter les effets de bords, on rend artificiellement impaires les trames de parole. 25
7. Procédé selon l'une des revendications précédentes, caractérisé par le fait que les ondelettes sont des ondelettes de Daubechies. 30
8. Procédé selon l'une des revendications précédentes, caractérisé par le fait que l'on utilise des ondelettes bi-orthogonales. 35
9. Procédé selon l'une des revendications précédentes, caractérisé par le fait qu'avant d'être stockés, les coefficients sont utilisés pour une synthèse d'évaluation (26), et ne sont stockés que lorsque la qualité de restitution est jugée satisfaisante. 40
10. Procédé selon l'une des revendications précédentes, caractérisé par le fait que le filtrage est fait par convolution. 45
11. Procédé selon l'une des revendications précédentes, caractérisé par le fait que pour une régularité supérieure à environ 16, le filtrage est fait par une multiplication dans l'espace fréquentiel dual. 50

55

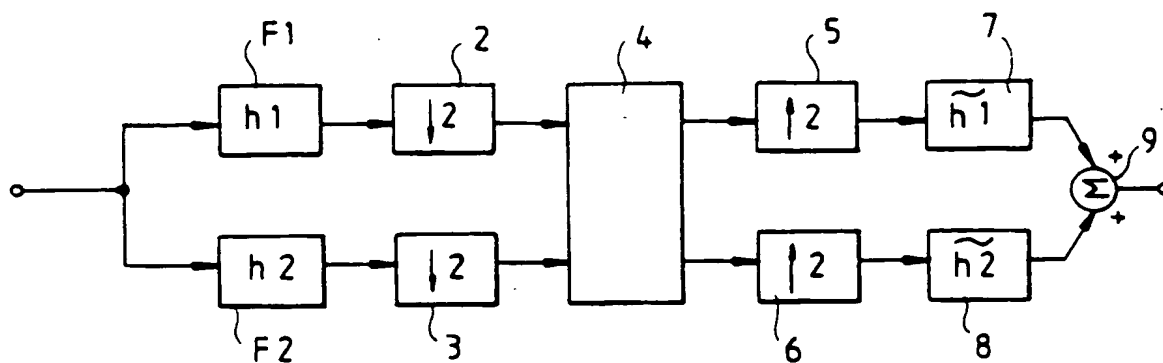


FIG. 1

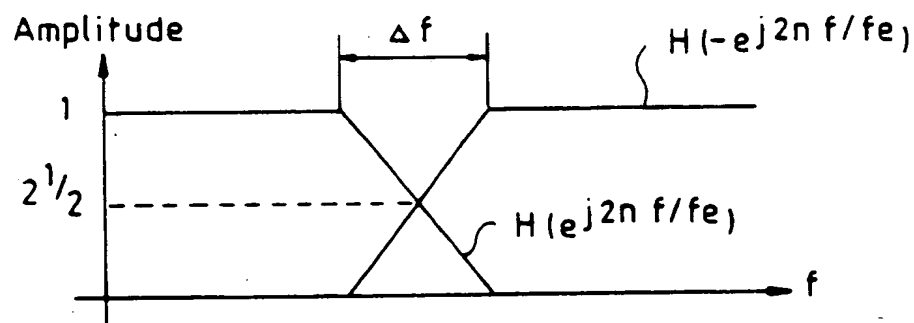


FIG. 2

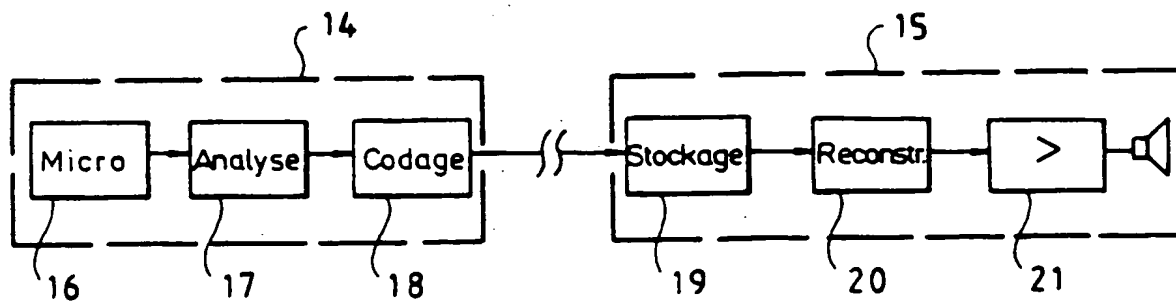
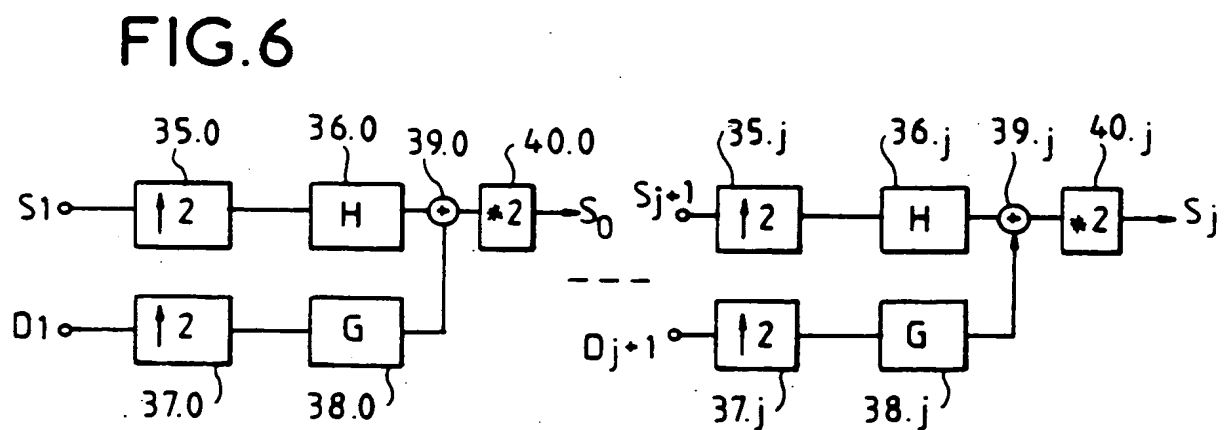
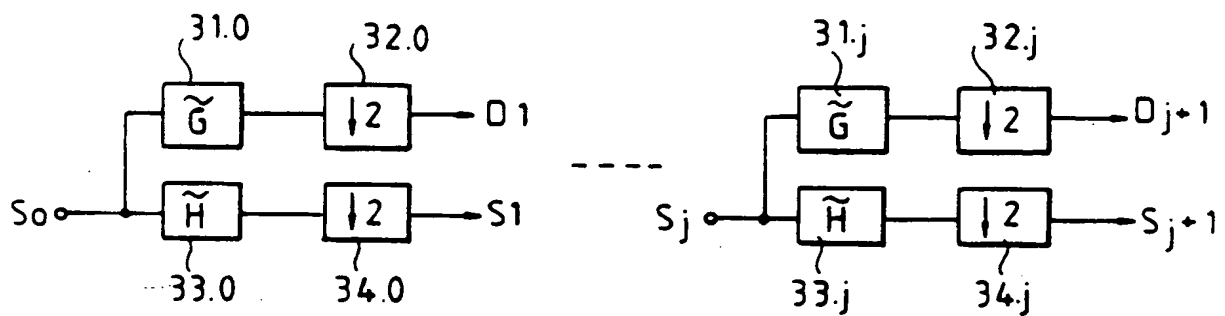
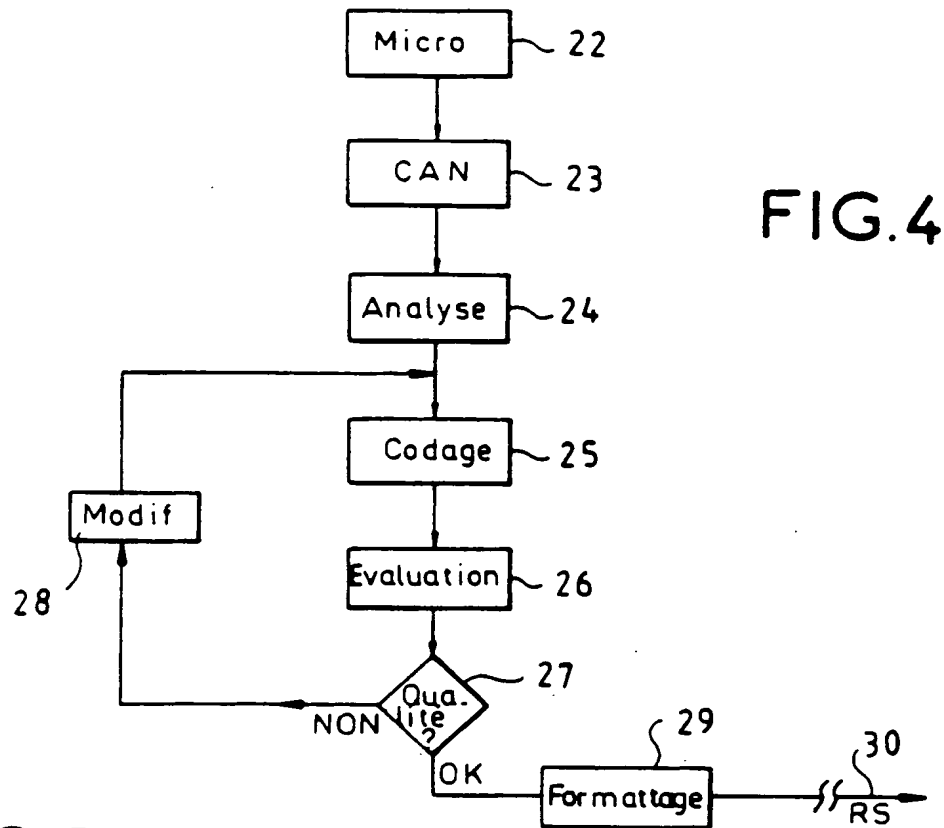


FIG. 3



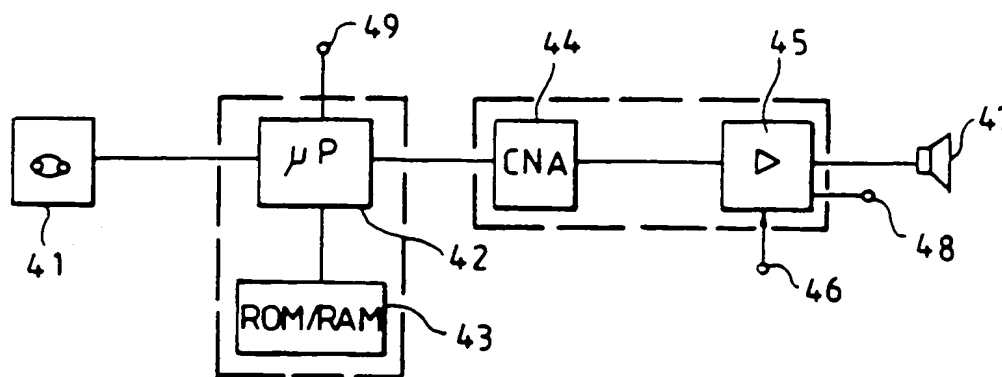


FIG. 7

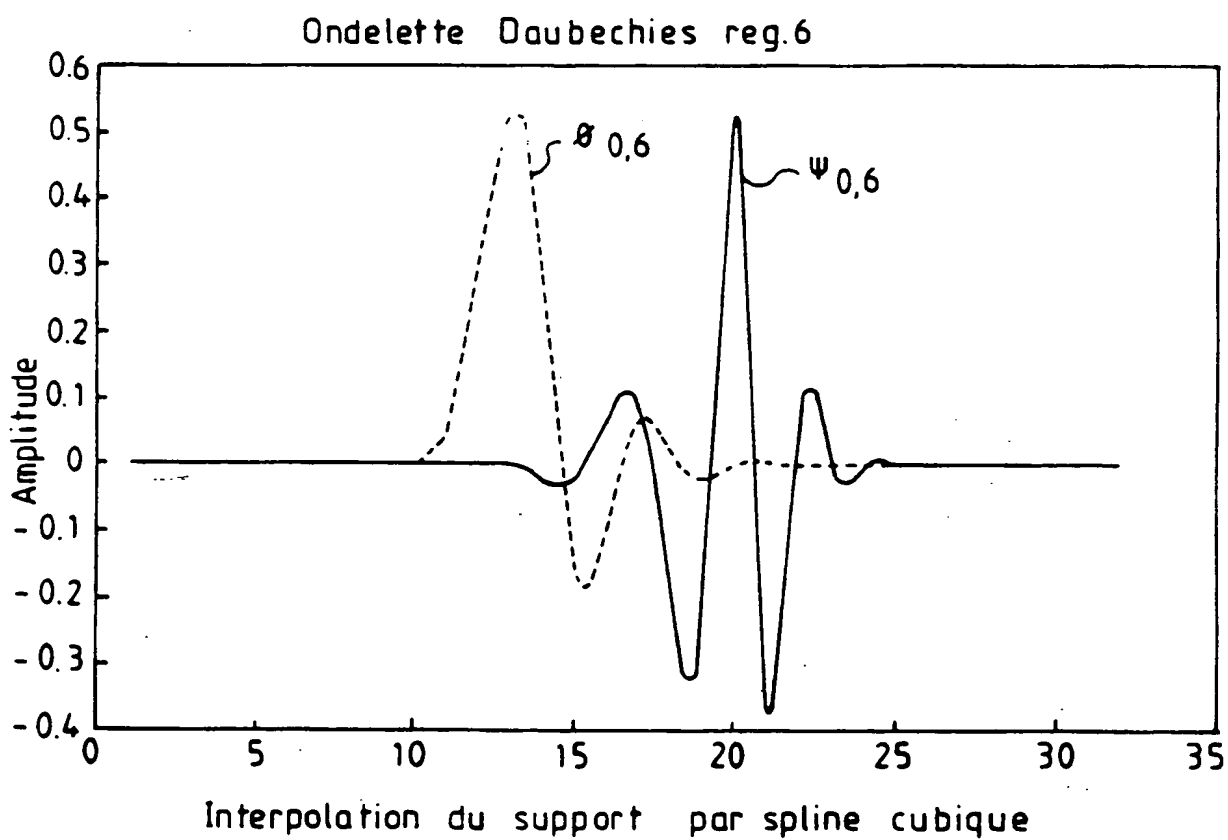


FIG. 8

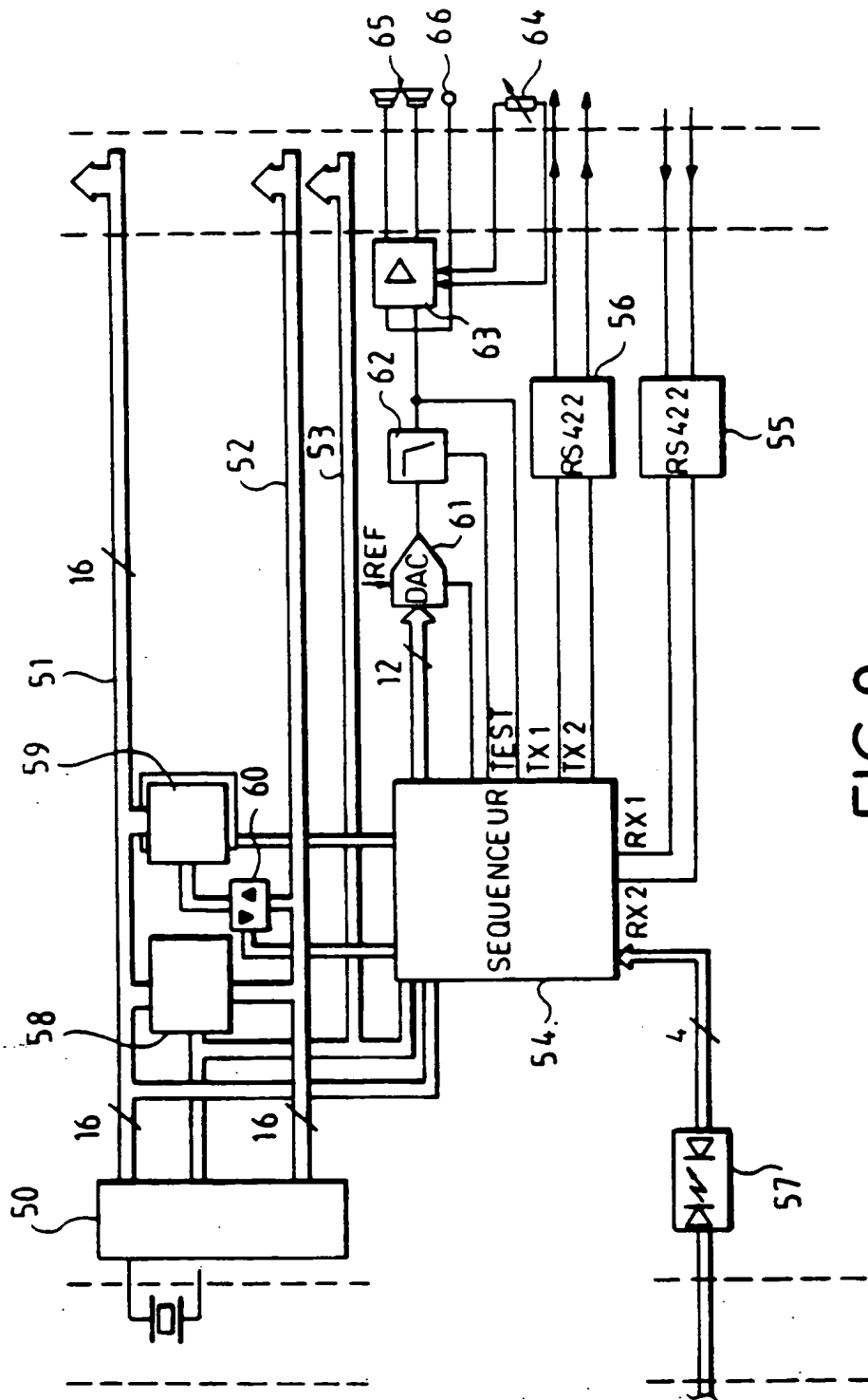


FIG. 9



Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numero de la demande

EP 92 40 1663

| DOCUMENTS CONSIDERES COMME PERTINENTS | | | |
|---|--|---|--|
| Catégorie | Citation du document avec indication, en cas de besoin, des parties pertinentes | Revendication concernée | CLASSEMENT DE LA DEMANDE (Int. Cl.5) |
| Y | COMPUTER MUSIC JOURNAL, vol. 12, no. 4, 1 janvier 1988, pages 11-20, Cambridge, Massachusetts; R. KRONLAND-MARTINET: "The wavelet transform for analysis, synthesis, and processing of speech and music sounds" * Page 13, colonne de gauche, lignes 3-23; pages 17-19: "Synthesis of sounds", "Modification of real sounds" * | 1,2,6-9 | G 10 L 5/02 G 10 L 9/16 |
| Y | COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, vol. XLI, 1988, pages 909-996; I. DAUBECHIES: "Orthonormal bases of compactly supported wavelets" * Page 965 - page 968, alinéa 4 * | 1,7-9 | |
| A | INTERNATIONAL JOURNAL ON PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE, vol. 1, no. 2, 1987, pages 273-302; R. KRONLAND-MARTINET et al.: "Analysis of sound patterns through wavelet transforms" * Page 280 - page 281, alinéa 3; page 290, alinéa 4.3 * | 1 | DOMAINES TECHNIQUES RECHERCHES (Int. Cl.5) G 10 L |
| Y | FR-A-2 648 567 (INSERM) * Page 1, lignes 3-7; page 3, ligne 34 - page 4, ligne 19; page 8, lignes 10-27; page 10, ligne 33 - page 11, ligne 31; page 13, ligne 6 - page 16, ligne 10 * | 2 | |
| A | --- -/- | 1,3-5 | |
| Le présent rapport a été établi pour toutes les revendications | | | |
| Lien de la recherche LA HAYE | | Date d'achèvement de la recherche 28-09-1992 | Examineur FARASSOPOULOS A. |
| CATEGORIE DES DOCUMENTS CITES | | T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant | |
| X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire | | | |

EPO FORM 1503 03.82 (P0402)

Office européen
des brevets

RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande

EP 92 40 1663

| DOCUMENTS CONSIDERES COMME PERTINENTS | | | |
|---|--|---|--------------------------------------|
| Catégorie | Citation du document avec indication, en cas de besoin, des parties pertinentes | Revendication concernée | CLASSEMENT DE LA DEMANDE (Int. Cl.5) |
| D,A | TRAITEMENT DU SIGNAL, vol. 7, no. 2, 1990, pages 101-115; P. MATHIEU et al.: "Compression d'image par transformée en ondelette et quantification vectorielle" * Pages 102-105; paragraphes 1,2.1,2.2,2.3,2.4 * | 1 | |
| Y | IDEM | 6 | |
| A | US-A-4 974 187 (LAWTON) * Colonne 1, ligne 1 - colonne 2, ligne 26; colonne 3, ligne 36 - colonne 5, ligne 50; colonne 13, lignes 9-14 * | 1,5 | |
| A | US-A-4 599 567 (GOUPILLAUD et al.) * Abrégé; colonne 13, lignes 10-37 * | 1 | |
| A | INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Albuquerque, New Mexico, 3-6 avril 1990, vol. 3, pages 1723-1726, IEEE, New York, US; M. VETTERLI et al.: "Wavelets and filter banks: relationships and new results" * Le document en entier * | 7,8,10, 11 | |
| A | INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Dallas, Texas, 6-9 avril 1987, vol. 2, pages 948-951, IEEE, New York, US; J.-S. LIENARD: "Speech analysis and reconstruction using short-time, elementary waveforms" * Abrégé; page 949, alinéa III; page 951, alinéa V * | 1 | |
| Le présent rapport a été établi pour toutes les revendications | | | |
| Lieu de la recherche LA HAYE | | Date d'achèvement de la recherche 28-09-1992 | Examinateur FARASSOPOULOS A. |
| <p>CATEGORIE DES DOCUMENTS CITES</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant</p> | | | |

EPO FORM 1503 (12.82) (P0402)